Universidade Estadual de Campinas
Faculdade de Engenharia Elétrica e de Computação

IA368 - Tópicos em Engenharia de Computação V

Prevendo o número de mortes por COVID-19 no mundo com inferência Bayesiana e modelos hierárquicos

Alunos:

147512, Nathália Menini Cardoso dos Santos

186604, Ricardo Gonçalves Molinari

216240, Eduardo Siqueroli

Professor:

Eduardo Valle

17 de julho de 2022

Conteúdo

1 Introdução

A doença COVID-19 surgiu através de um novo tipo de coronavírus, conhecido como SARS-CoV-2 (*Severe Acute Respiratory Syndrome Coronavirus-2*). O SARS-CoV-2 difere-se de outros coronavírus que, em sua imensa maioria das vezes, espalham-se entre seres humanos causando apenas um resfriado comum. A COVID-19 é uma doença infecciosa respiratória aguda que é transmitida, principalmente, através do trato respiratório [1, 2].

É amplamente difundido que em uma parcela das pessoas infectadas, a doença poderá evoluir de modo desfavorável e, inclusive, levar a óbito. A taxa de mortalidade é maior entre os idosos e entre outras etnias que não os caucasianos. Além disso, uma taxa de mortalidade nitidamente mais alta foi observada para pessoas portadoras de comorbidades pré-existentes como, por exemplo, obesidade, diabetes mellitus não controlada, malignidades no ano anterior, doenças renais, doenças respiratórias crônicas que não asma, doença hepática crônica, acidente vascular cerebral, demência, transplante de órgãos e imunossupressão [3].

A falta de preparação para combater o surto da doença, juntamente com a baixa compreensão do agente causador, faz com que o problema seja ainda mais complexo e difícil de solucionar [4]. Dessa maneira, é de extrema importância entender a dinâmica da pandemia, para que medidas públicas possam ser tomadas no tempo adequado e de maneira apropriada. Propomos, com esse trabalho, utilizar dados globais sobre a pandemia para melhor elucidar os fatores que contribuem para uma maior (ou menor) probabilidade de evoluir para óbito ao contrair a doença COVID-19.

2 Conjunto de dados

Os dados utilizados para este trabalho foram extraídos do repositório do github do *Our World in Data*[1]. Desse conjunto de dados, utilizamos as seguintes informações:

- **Número de casos e mortes confirmadas.** Os dados são provenientes do repositório CSSE (*Systems Science and Engineering*) da Universidade Johns Hopkins.
- **Nome do país.** Informação tanto de localização geográfica quanto do código universal do país.
- **Índice de Desenvolvimento Humano (IDH).** Índice composto que mede o desempenho médio em três dimensões básicas do desenvolvimento humano – uma vida longa

[1]Extraído dia 10/06/2022 às 16:47 de https://github.com/owid/covid-19-data/tree/master/public/data

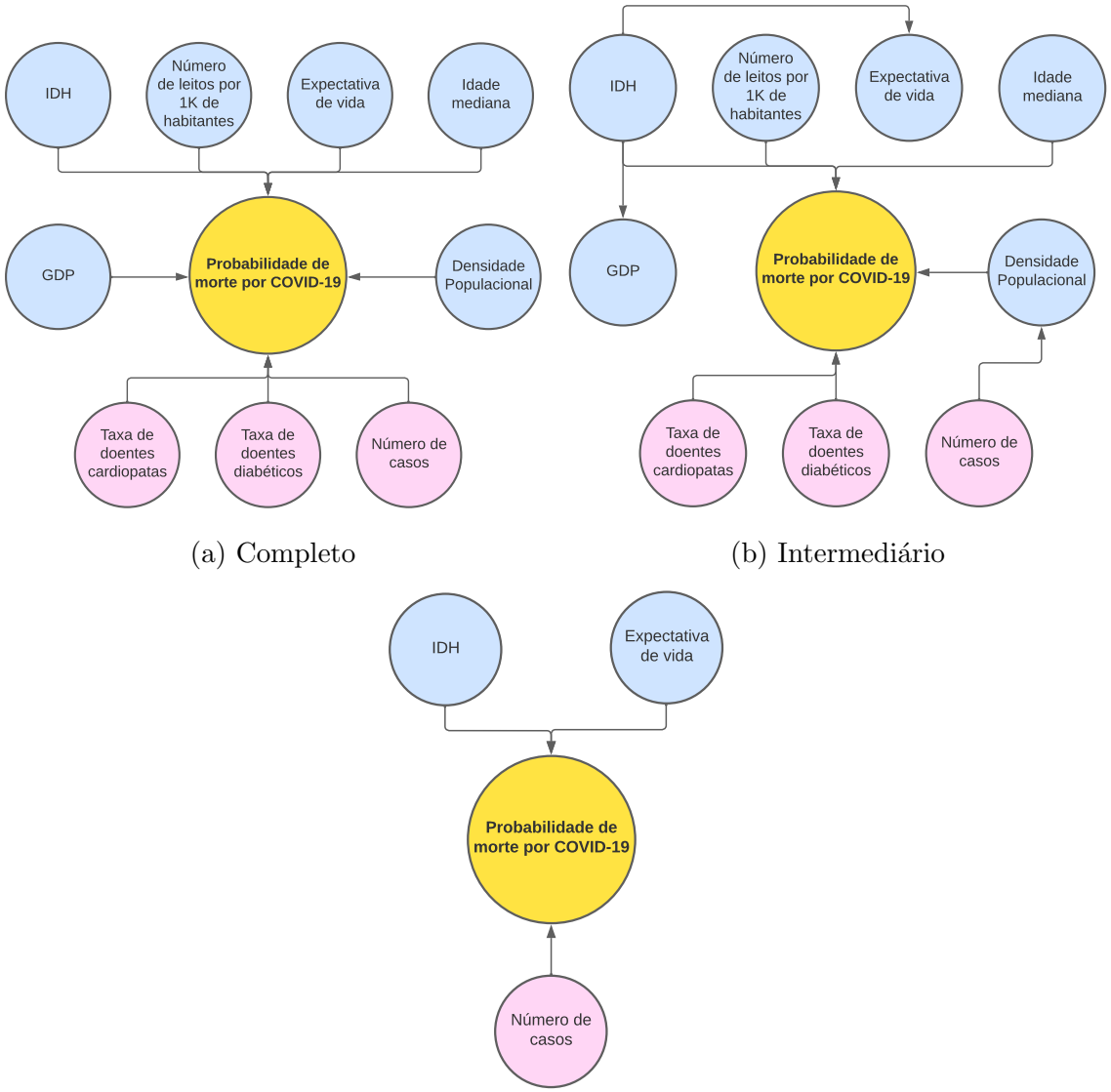
e saudável, conhecimento, economia interna e um padrão de vida decente. Valores para 2019, fornecido pelo UNDP (*United Nations Development Programme*).

- **Número de leitos de hospital por 1.000 habitantes.** Informação mais recente disponível por país, extraído de OECD, Eurostat, World Bank, *national government records* e outras fontes.
- **Expectativa de vida.** Expectativa de vida ao nascer em 2019. Extraído de Clio Infra e *United Nations Population Division*.
- **Idade mediana.** Idade média da população, projeção da Organização da Nações Unidas (ONU) para 2020.
- **Taxa de mortes por cardiopatia.** Taxa de mortalidade por doença cardiovascular em 2017 (número anual de mortes por 100.000 pessoas). Extraído de *Global Burden of Disease Collaborative Network*.
- **Taxa de prevalência de diabetes.** Prevalência de diabetes (% da população de 20 a 79 anos) em 2017. Extraído de *World Bank World Development Indicators*.
- **Densidade populacional.** Número de pessoas dividido por área de terra, medido em quilômetros quadrados.
- ***Gross domestic product* (GDP).** Produto interno bruto em paridade de poder de compra.

3 Modelo Causal

Com base nas informações disponíveis apresentadas na seção anterior, realizamos uma análise preliminar para entendimento de um possível modelo causal que poderia ser empregado neste projeto. Com isso, foi possível selecionar as variáveis preditoras, as variáveis de controle e a variável de interesse.

Mais especificamente, na Fig. 1 temos que as variáveis preditoras estão destacadas com a cor azul, as variáveis de controle com a cor rosa e a variável de interesse com a cor amarela. Considerando uma possível relação de causalidade entre as variáveis preditoras e/ou de controle, optamos por testar diferentes configurações de modelo causal: Modelo Completo, contém todas as variáveis diretamente se relacionando com a variável de interesse (Fig. 1a); Modelo Intermediário, existem algumas relações de causalidade entre as covariáveis e variáveis de controle (1b) e; Modelo Reduzido, versão reduzida que considera apenas algumas das variáveis (Fig. 1c).



(a) Completo

(b) Intermediário

(c) Reduzido

Figura 1: Modelos de causalidade.

- **IDH e GDP.** Em teoria, quanto maior o valor dessas variáveis, mais desenvolvido e melhor é a qualidade de vida do país é, e isso poderia levar a uma menor probabilidade de morte.

- **Número de leitos de hospital por 1.000 habitantes.** Quanto menor a quantidade de leitos disponível, mais pessoas ficariam sem acesso a um melhor tratamento, levando a uma maior a probabilidade de morte.

- **Expectativa de vida.** Quanto maior a expectativa de vida, é razoável inferir que existam mais idosos e, portanto, poderia elevar a probabilidade de morte.

- **Idade mediana.** Quando maior a mediana da idade, podemos inferir que trata-se de uma população mais envelhecida e, com isso, teríamos uma probabilidade de morte maior.

- **Densidade populacional.** Quando maior a densidade populacional, mais propensa uma pessoa poderia ser de se contagiar e isso aumentaria a probabilidade de morte.

Embora as variáveis de controle (taxa de mortes por cardiopatia de controle, taxa de prevalência de diabetes e número de casos) não sejam variáveis cujo efeito direto seja de interesse neste projeto, optamos por considerá-las com a finalidade de controlar o efeito que possam ter na probabilidade de morte por COVID-19. Acredita-se que um país com uma maior taxa de pessoas com comorbidades pré-existentes possa levar a uma maior probabilidade de morte, assim como um maior número de casos também tende a aumentar a quantidade de mortes.

4 Modelo Bayesiano

4.1 Verossimilhança

Neste estudo, consideramos o número acumulado de casos e óbitos de 159 países desde o início da pandemia em cada país até 09 de junho de 2022. A abordagem proposta baseia-se em um modelo de regressão binomial para a resposta Y (o número total de mortes por COVID-19 em cada país), na presença das seguintes variáveis explicativas: número de leitos de hospital por 1.000 habitantes, expectativa de vida, idade mediana, taxa de mortes por cardiopatia, taxa de prevalência de diabetes e número de casos diagnosticados de COVID-19. Sendo razoável supor que $Y_{ji} \sim Binomial(n_{ji}, \theta_{ji})$, a Eq. 1 descreve a probabilidade de se observar uma morte por COVID-19 do i-ésimo país e do j-ésimo nível do IDH, dado o tamanho da população n_{ji} e a proporção θ_{ji}.

$$P(Y_{ji} = y_{ji}|\theta_{ji}, n_{ji}) = \binom{n_{ji}}{y_{ji}} \theta_{ji}^{y_{ji}} (1 - \theta_{ji})^{n_{ji} - y_{ji}} \tag{1}$$

4.2 Regressão logística

Tomemos como exemplo o modelo de causalidade ilustrado na Fig. 1a. Como estamos interessados em analisar a relação das variáveis explicativas com a probabilidade de morte, podemos utilizar uma transformação de θ_{ji}, como descrita na Eq. 2, onde x_{j1}: é uma valor constante igual a 1 (referente ao intercepto); x_{j2} é o número total de casos de Covid-19; x_{j3} é o número de leitos de hospital por 1.000 habitantes; x_{j4} é a expectativa de vida; x_{j5} é a idade mediana; x_{j6} é a taxa de prevalência de diabetes; x_{j7} é a taxa de mortes por cardiopatia de controle; x_{j8} é a densidade populacional; x_{j9} é o GDP. O índice i representa o i-ésimo país e j representa o

j-ésimo nível hierárquico. Ainda, os termos β_{ji} e x_{ji} podem ser escritos como o produto interno das suas formas matriciais como $\beta_j = (\beta_{j1}, \beta_{j2}, ..., \beta_{j9})^T$ e $x_j = (1, x_{j2}, x_{j3}, ..., x_{j9})^T$. Todas as variáveis x_j passaram por uma padronização dos seus valores para média igual a zero e variância unitária, ou seja, $x_i = (x_i - \mathbb{E}(x))/Var(x)$.

$$\theta_{ji} = logit^{-1}(\sum_{i=1}^{9} \beta_{ji} x_{ji}) \tag{2}$$

4.3 Modelo hierárquico

O modelo hierárquico foi estruturado supondo que o IDH possa ter forte influência na probabilidade de morte por COVID-19 e, por isso, optamos por categorizar essa variável em três níveis: baixo, médio e alto. Dado a versão categorizada do IDH, propomos uma modelagem hierárquica, em que os níveis da hierarquia se dariam pelos níveis do IDH. Dessa forma, embora as probabilidades de morte por COVID-19 sejam eventos gerados por uma mesma distribuição de probabilidade, possuem valores distintos entre níveis de hierarquia. Dessa forma, os parâmetros β_{ji} que fazem referência a uma mesma variável x_{ji} em diferentes níveis hierárquicos tem origem em uma mesma distribuição normal, com valor médio μ_i e variância σ_i. Ainda considerando o exemplo do modelo de causalidade ilustrado na Fig. 1a, os hiperparâmetros podem ser descritos pelos conjuntos $\mu = (\mu_1, \mu_2, ..., \mu_9)^T$ e $\sigma = (\sigma_1, \sigma_2, ..., \sigma_9)^T$. Nesse sentido, a probabilidade de um dado conjunto de parâmetros B_j dado os hiperparâmetros μ e σ pode ser descrito como na Eq. 3.

$$\begin{gathered} \beta_{ji} \sim \mathcal{N}(\mu_i, \sigma_i) \\ P(B_j = \beta_j | \mu, \sigma) \propto \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{\beta_j - \mu}{\sigma}\right)^2\right) \end{gathered} \tag{3}$$

4.4 Distribuição conjunta a priori dos hiperparâmetros

O modelo de distribuição a priori dos hiperparâmetros μ foram modelados como similares a distribuição normal, com valor médio igual a zero e variância igual a cinco, enquanto que σ os hiperparâmetros σ foram modelados como similares a distribuição exponencial, com $\lambda = 5$. Em todos os modelos de causalidade foi utilizado a mesma distribuição a priori dos hiperparâmetros. Ainda, as distribuições a priori dos hiperparâmetros μ e σ foram considerados independentes, resultando em uma distribuição conjunta a priori com probabilidade descrita pela Eq. 4.

$$
\begin{aligned}
P(\mu) &\sim \mathcal{N}(0, 5) \\
P(\sigma) &\sim Exp(5) \\
P(\mu, \sigma) &\propto P(\mu)P(\sigma)
\end{aligned} \tag{4}
$$

Os valores dos hiperparâmetros da distribuição a priori foram definidos a partir da exploração das predições a priori em comparação com os dados, de forma que a massa de probabilidade das predições englobasse os dados observados. O cálculo das predições a priori do número de mortos por COVID-19 foram estimadas utilizando-se somente as distribuições a priori de μ e σ, sem considerar os dados utilizados para o cálculo da distribuição conjunta a posteriori.

4.5 Distribuição conjunta a posteriori dos parâmetros e hiperparametros

O modelo final é caracterizado pelos parâmetros β e hiperparâmetros μ e σ, cuja distribuição conjunta a posteriori pode ser descrita pela Eq. 5, onde: y_{ji}, n_{ji}, x_{ji} representam, respectivamente, a quantidade de mortes por COVID-19, tamanho da população e covariáveis observadas do i-ésimo país do j-ésimo nível hierárquico do IDH e; K_j representa a quantidade de países em cada nível hierárquico e $\beta_j = (\beta_{j1}, \beta_{j2}, ..., \beta_{j9})$, para $j = \{1, 2, 3\}$, representando o vetor dos parâmetros da regressão para cada nível hierárquico do IDH (baixo, médio e alto).

$$
\begin{aligned}
P(\beta, \mu, \sigma | y, n, x) &\propto P(\mu, \sigma)P(\beta|\mu, \sigma)P(y|\beta, n, x) \\
&\propto P(\mu, \sigma) \prod_{j=1}^{3} P(\beta_j|\mu, \sigma) \prod_{i=1}^{K_j} P(y_{ji}|\beta_j, n_{ji}, x_{ji})
\end{aligned} \tag{5}
$$

A probabilidades da distribuição conjunta a posteriori foi calculada com o algoritmo Markov Chain Monte Carlo (MCMC) [5], implementado com a linguagem de programação R na plataforma Stan [6].

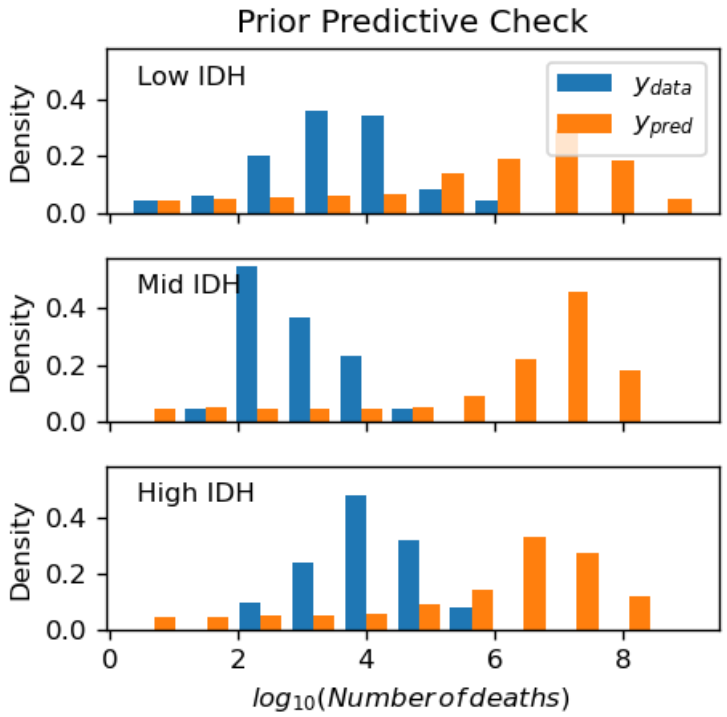


Figura 2: *Prior predictive check para o Modelo Completo.*

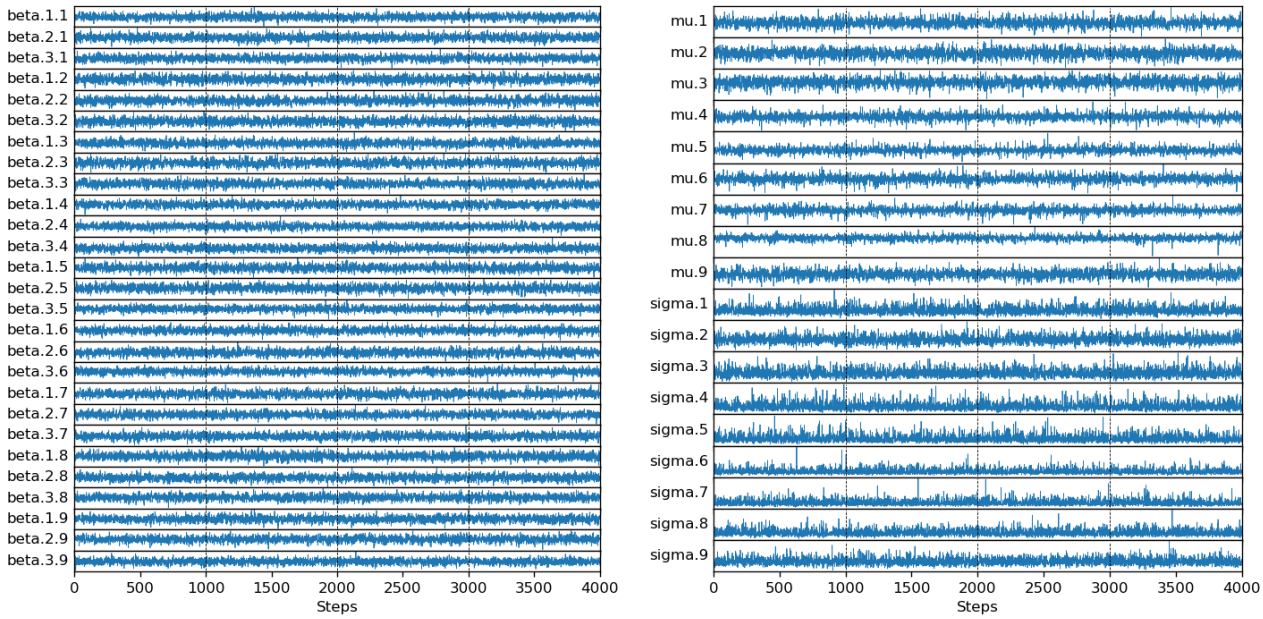
5 Resultados

5.1 Predição do número de mortes por COVID-19 a partir da distribuição a priori

A predição do número de mortes por COVID-19 a partir da distribuição a priori em comparação com os dados pode ser visualizada na Fig. 2, sugerindo um ajuste aceitável, mas não suficiente, do modelo aos dados. Sendo assim, é necessário considerar as informações presentes nos dados para se obter uma inferência mais precisa.

5.2 Modelo Completo

No chamado Modelo Completo (Fig. 1a) foi utilizado todas as variáveis disponíveis no nosso conjunto de dados. Dessa maneira, utilizamos a quantidade total de casos (β_2), número de leitos (β_3), expectativa de vida (β_4), idade mediana (β_5), taxa de prevalência de diabetes (β_6), taxa de mortes por cardiopatia (β_7), densidade populacional (β_8) e GDP (β_9). Na Fig. 3 apresentamos as 4 cadeias geradas pelo algoritmo MCMC, após realizar para todos os parâmetros e hiperparâmetros o *burn* inicial de 50% das amostras, resultando em um conjunto total de 4.000 amostras. A inspeção visual do valor médio e da variância das cadeias sugere que os parâmetros



(a) Parâmetros

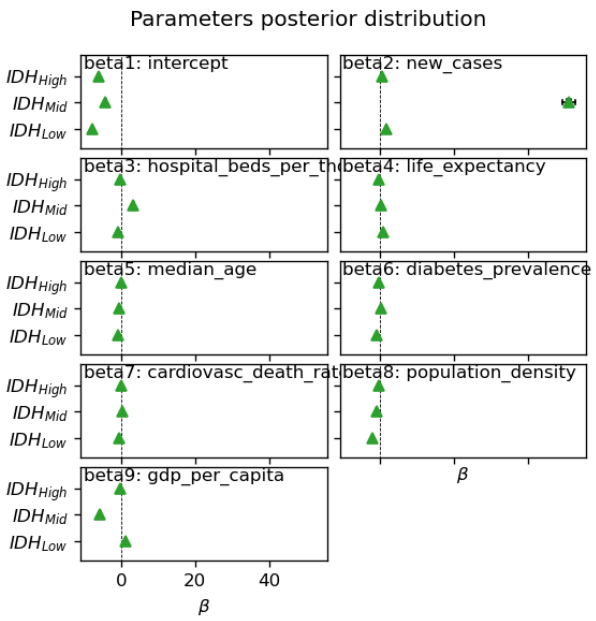
(b) Hiperparâmetros

Figura 3: 4 *chains* geradas pelo HMC (após o *burn* inicial) para o Modelo Completo. Cada uma das *chains* estão separadas pelas retas tracejadas verticais.

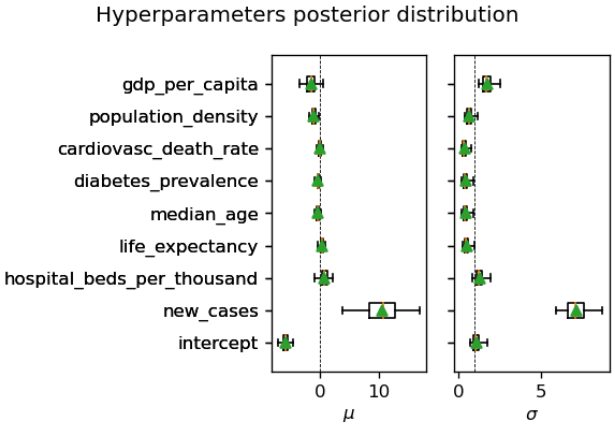
e os hiperparâmetros convergiram para o mesmo resultado.

Na Fig. 4a apresentamos as distribuições para β e na Fig. 4b para os hiperparâmetros μ e σ do modelo. É possível reconhecer que a variável número de casos (indicada no gráfico como `new_cases`) tem grande eficiência na explicabilidade do modelo para o IDH médio, de modo que, quanto maior a quantidade de casos, maior será a probabilidade e quantidade de mortes - para os demais grupos de IDH, essa variável parece não ser tão decisiva. Já em relação o GDP (β_9), vemos que para o nível médio de IDH, quanto maior é o valor dessa variável, menor é a probabilidade de morte. Entretanto, as demais variáveis do modelo não demonstraram tanta significância comparadas com as citadas acima. A análise dos hiperparâmetros (Fig. 4b) vai ao encontro da da análise dos parâmetros, sugerindo o que o número total de casos e o GDP são variáveis importantes para descrever o modelo.

Por fim, na Fig. 5 apresentamos a verificação da predição a posteriori do modelo (*posterior predictive check*). Mais especificamente, na Fig. 5a, temos o *posterior predictive check* do modelo de forma geral (sem agrupar pelo nível da hierarquia), e em 5b separado pelos níveis do IDH. Em ambos os cenários, vemos que as distribuições geradas se assemelham muito às distribuições presentes nos dados, indicando que o modelo apresenta um bom ajuste, sendo capaz de gerar valores muito próximos aos valores presentes nos dados amostrados.

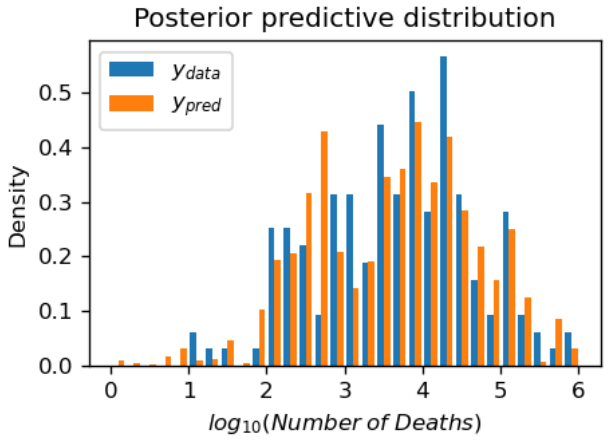


(a) Parâmetros

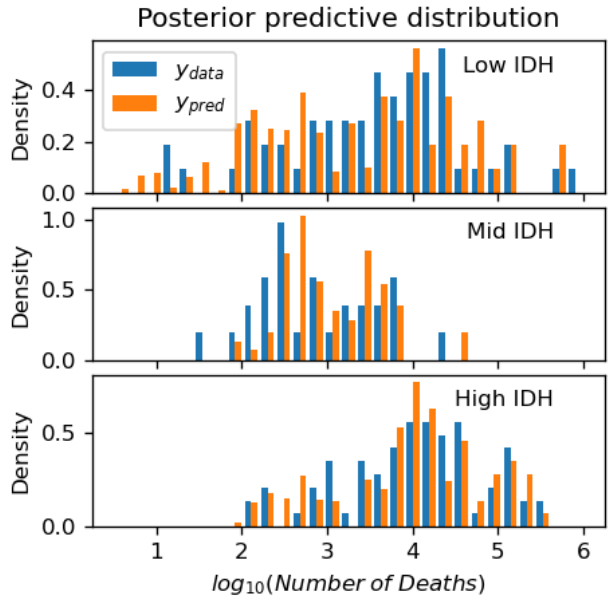


(b) Hiperparâmetros

Figura 4: Distribuição dos parâmetros e dos hiperparâmetros para o Modelo Completo.



(a) Geral

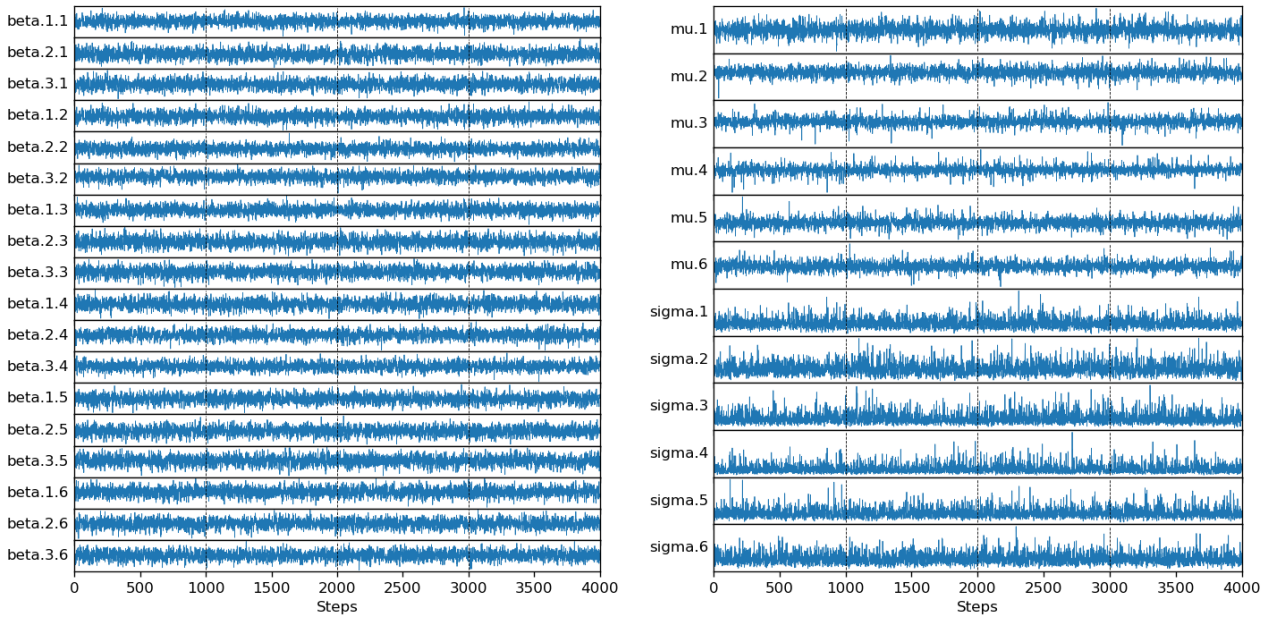


(b) Por grupo de IDH

Figura 5: *Posterior predictive check* para o Modelo Completo.

5.3 Modelo Intermediário

Em um segundo momento, consideramos o modelo causal da Fig. 1b. Nesse modelo, consideramos as seguintes variáveis: número de leitos (β_2), idade mediana (β_3), taxa de prevalência de diabetes (β_4), taxa de mortes por cardiopatia (β_5), densidade populacional (β_6). Ou seja,



(a) Parâmetros

(b) Hiperparâmetros

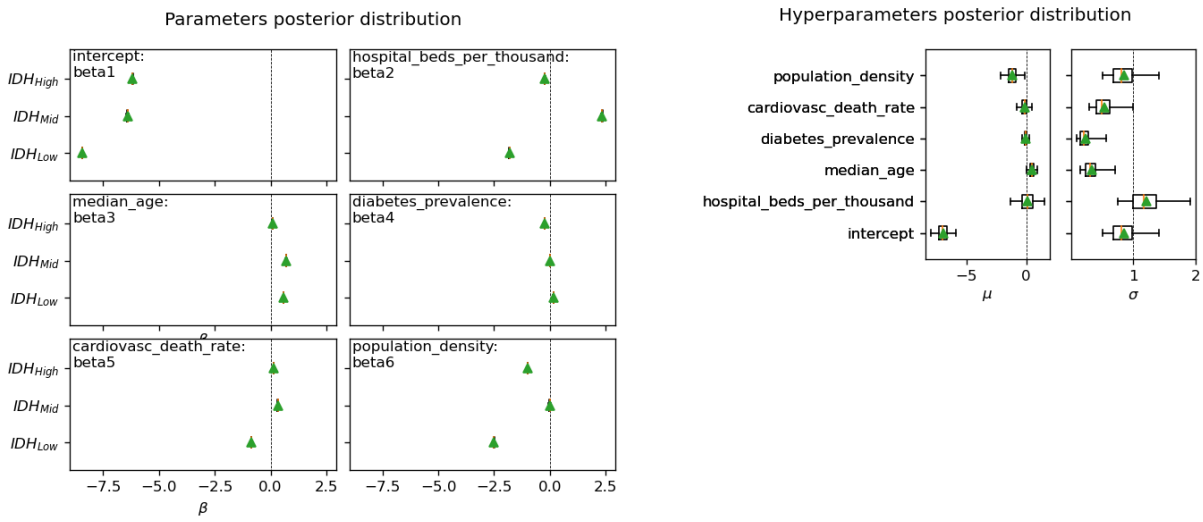
Figura 6: 4 *chains* geradas pelo HMC (após o *burn* inicial) para o Modelo Intermediário. Cada uma das *chains* estão separadas pelas retas tracejadas verticais.

em relação ao modelo da Seção 5.2, removemos as variáveis GDP e expectativa de vida, por possuírem uma relação de causalidade com o IDH. Também removemos o número total de casos, pela sua possível relação com a variável de densidade populacional.

De forma similar ao que foi observado no modelo da Seção 5.2, na Fig. 6 apresentamos as 4 cadeias geradas pelo algoritmo MCMC, após realizar para todos os parâmetros e hiperparâmetros o *burn* inicial de 50% das amostras, resultando em um conjunto total de 4.000 amostras. A inspeção visual do valor médio e da variância das cadeias sugere que os parâmetros e os hiperparâmetros convergiram para o mesmo resultado.

Na Fig. 7 exibimos a distribuição dos parâmetros e dos hiperparâmetros para o Modelo Intermediário. Em 7a, é possível ver que as variáveis idade mediana, prevalência de diabetes e taxa de morte por doença cardiovascular não aparentam possuir poder de explicabilidade em relação a variável de interesse. Já em relação a variável que mede a quantidade de leitos, vemos que para países com IDH baixo, quanto maior é a quantidade de leitos, menor é a probabilidade de morte, ocorrendo o oposto com países com IDH médio - para IDH alto, a variável não se mostrou significativa. Inesperadamente, observou-se que os países com IDH alto e IDH baixo, quanto maior a densidade populacional, menor a probabilidade de morte, sendo o contrário do que se esperaria para essa variável, já que uma maior densidade populacional expõe mais a população a disseminação do vírus.

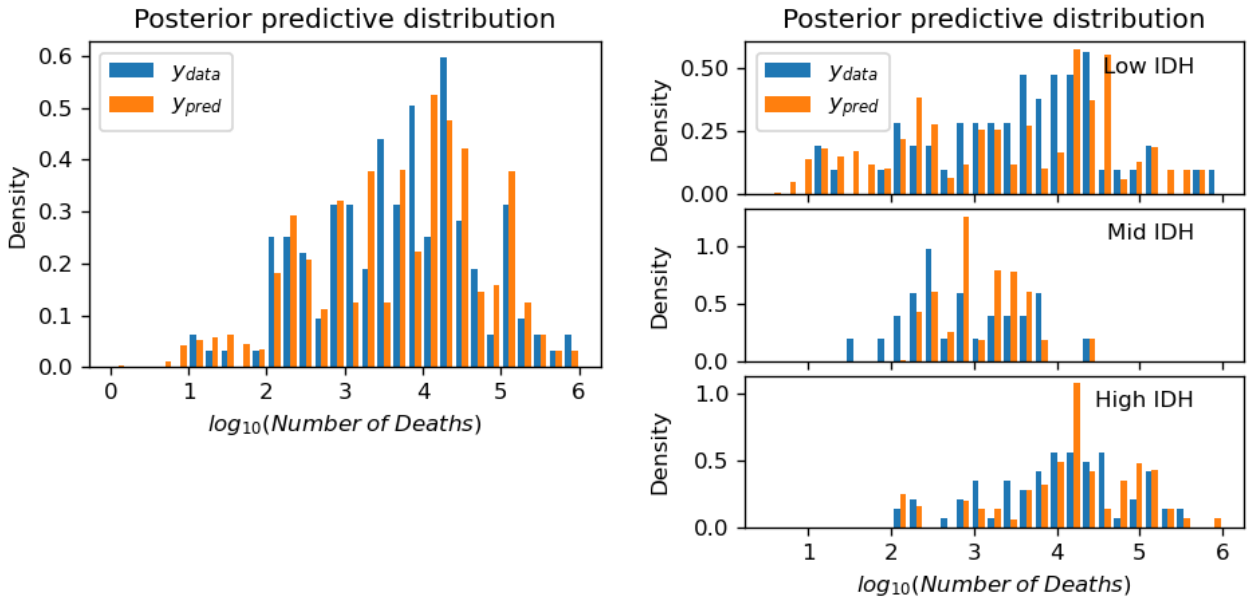
Por fim, na Fig. 8 apresentamos o *posterior predictive check*. Mais especificamente,



(a) Parâmetros

(b) Hiperparâmetros

Figura 7: Distribuição dos parâmetros e dos hiperparâmetros para o Modelo Intermediário.

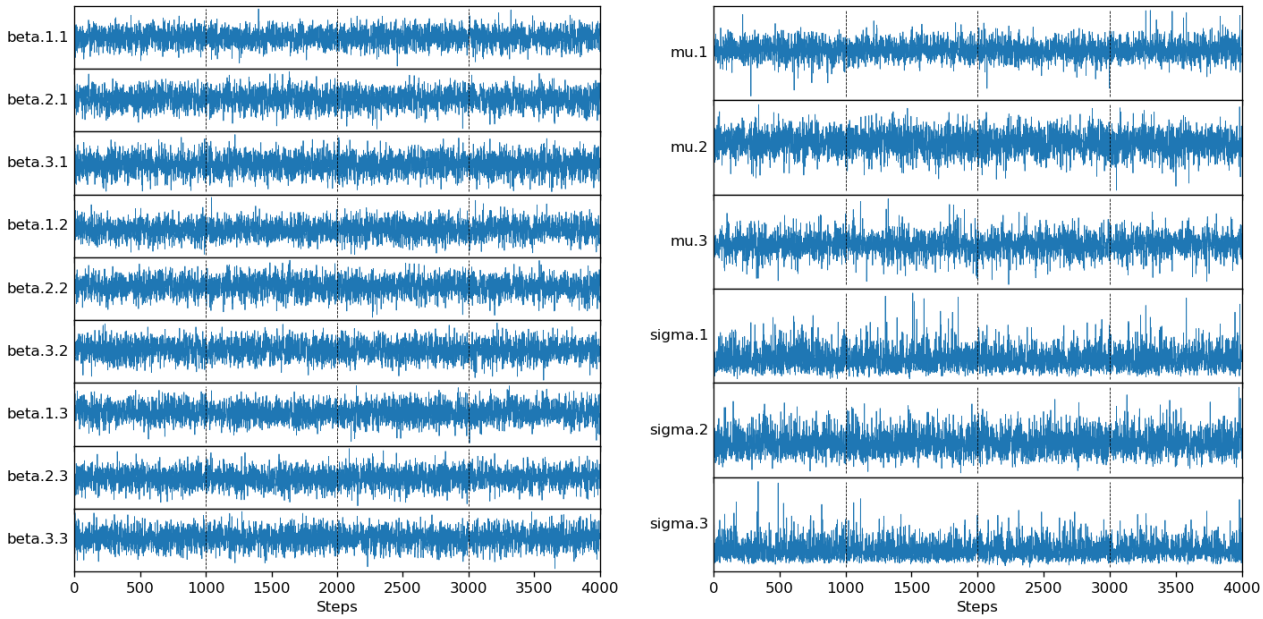


(a) Geral

(b) Por grupo de IDH

Figura 8: *Posterior predictive check* para o Modelo Intermediário.

em 8a temos o *posterior predictive check* do modelo de forma geral (sem agrupar pelo nível da hierarquia), e em 8b separado pelos níveis do IDH. Em ambos os cenários, vemos que as distribuições geradas se assemelham muito às distribuições presentes nos dados, indicando que o modelo apresenta um ajuste aceitável, sendo capaz de gerar valores muito próximos aos valores presentes nos dados amostrados. Vale ressaltar que as predições feitas para o grupo com IDH intermediário apresentaram valores mínimos com uma magnitude de diferença dos dados, prevendo um mínimo em torno de 200 mortes enquanto que os dados sugerem 20 mortes.



(a) Parâmetros

(b) Hiperparâmetros

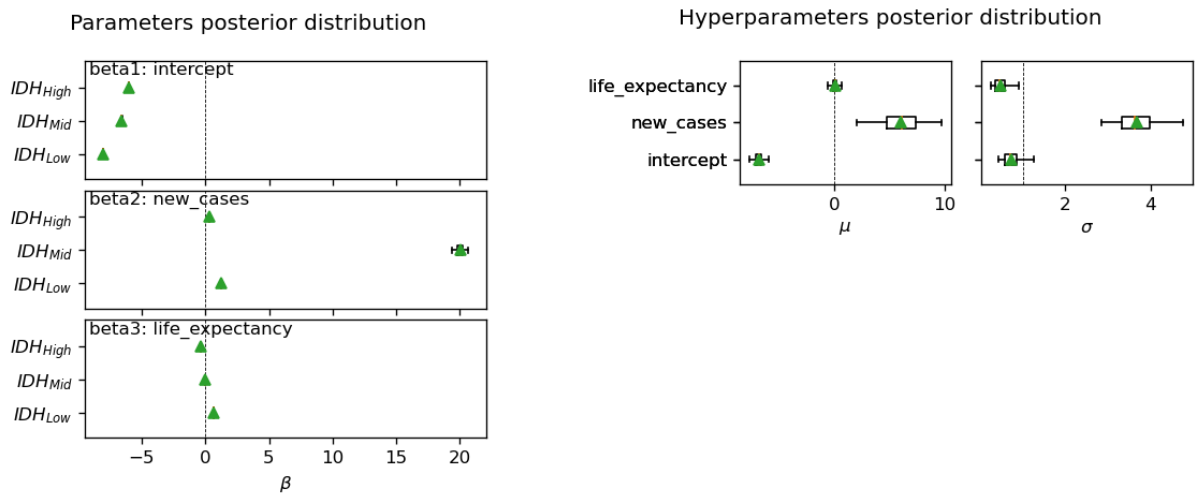
Figura 9: 4 *chains* geradas pelo HMC (após o *burn* inicial) para o Modelo Reduzido. Cada uma das *chains* estão separadas pelas retas tracejadas verticais.

5.4 Modelo Reduzido

Para finalizar as análises referentes a este problema, consideramos o Modelo Reduzido, ilustrado pelo modelo causal da Fig. 1c. Nesse modelo, levamos em consideração as seguintes variáveis: número total de casos (β_2) e expectativa de vida (β_3). De forma similar ao que foi observado nos modelos das Seções 5.2 e 5.3, na Fig. 9 apresentamos as 4 cadeias geradas pelo algoritmo MCMC, após realizar para todos os parâmetros e hiperparâmetros o *burn* inicial de 50% das amostras, resultando em um conjunto total de 4.000 amostras. A inspeção visual do valor médio e da variância das cadeias sugere que os parâmetros e os hiperparâmetros convergiram para o mesmo resultado.

Na Fig. 10 exibimos a distribuição dos parâmetros e dos hiperparâmetros para o Modelo Reduzido. Em 10a, é possível ver que a variável expectativa de vida (β_3) não aparenta possuir poder de explicabilidade em relação a variável de interesse no nível de IDH médio, enquanto no nível alto e baixo apresentam uma pequena influencia. Entretanto, a variável número total de casos (β_2:`new_cases` aparenta ter grande influencia em todos o níveis de hierarquia do modelo. Isso é possível graças a relação de β_2 com a quantidade de mortes, já que tem relação direta com a variável de interesse.

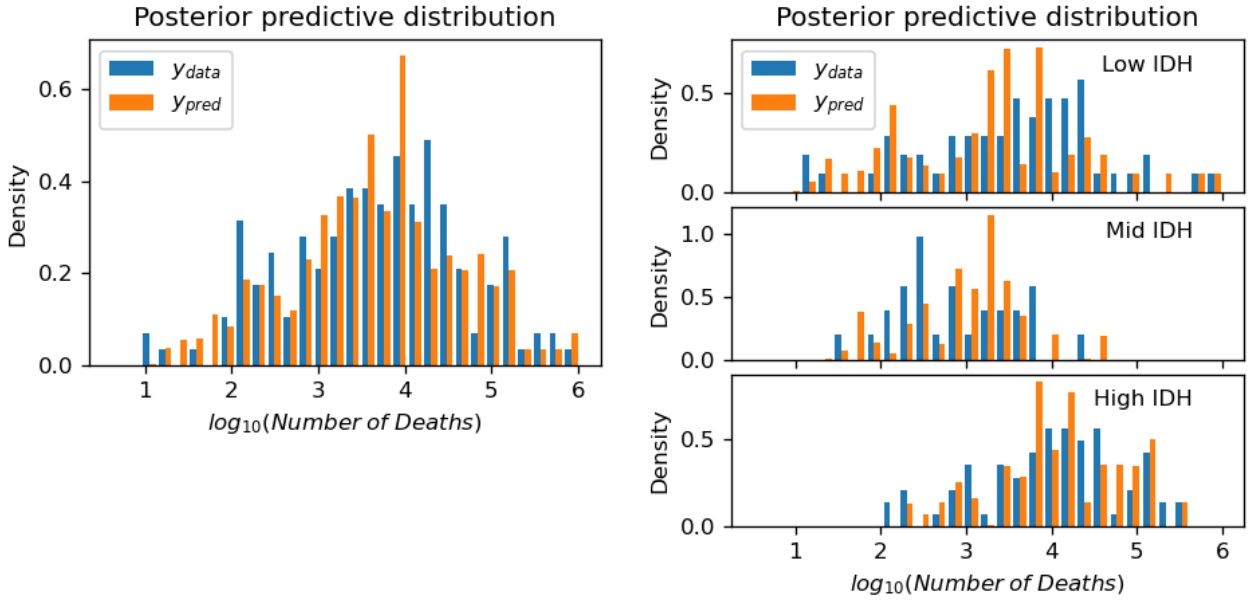
Por fim, na Fig. 11 apresentamos o *posterior predictive check*. Mais especificamente, em 11a temos o *posterior predictive check* do modelo de forma geral (sem agrupar pelo nível



(a) Parâmetros

(b) Hiperparâmetros

Figura 10: Distribuição dos parâmetros e dos hiperparâmetros para o Modelo Reduzido.



(a) Geral

(b) Por grupo de IDH

Figura 11: *Posterior predictive check* para o Modelo Reduzido.

da hierarquia), e na Fig. 11b separado pelos níveis do IDH. Como nos outros modelos ambos os cenários, as distribuições geradas se assemelham muito às distribuições presentes nos dados, indicando que o modelo também apresenta um ajuste aceitável mesmo se tratando de um modelo mais reduzido, sendo capaz de gerar valores muito próximos aos valores presentes nos dados amostrados.

6 Etapas do Desenvolvimento

Desenvolvemos este projeto em 4 grandes tarefas, além da análise final de resultados. São elas: (1) definição do problema e do conjunto de dados, (2) desenho do modelo causal, (3) entendimento e definição do modelo analítico e (4) implementação do modelo computacional. Abaixo listamos as principais decisões e dificuldades encontradas em cada uma das tarefas:

1. **Definição do problema e conjunto de dados.** Para a definição do problema, decidimos escolher um tema muito presente nos dias atuais. Dessa maneira, isso facilitaria a pesquisa pelo conjunto de dados.

2. **Modelo causal.** Uma vez que encontramos o conjunto de dados e as variáveis presentes, desenhamos alguns modelos causais. A maior dificuldade foi pensar nas possíveis relações de causalidade entre as variáveis e como isso poderia interferir na modelagem analítica.

3. **Modelo analítico.** Em um primeiro momento, despendemos um tempo considerável para um completo entendimento da modelagem hierárquica, em conjunto com um modelo de regressão. Após muito estudo e auxílio do professor, foi possível compreender completamente o modelo analítico proposto e apresentado neste trabalho.

4. **Modelo computacional.** A principal dificuldade encontrada nesse projeto está relacionada a essa etapa. Diversas tentativas de execução do algoritmo pystan foi tentada, utilizando o Python como linguagem de programação. Após exaustivas tentativas e insucessos, decidimos mudar de linguagem de progração e utilizar o RStan, no R. Após essa troca de linguagem, a modelagem prosseguiu sem muitos problemas.

6.1 Futuros trabalhos

Dada as dificuldades encontradas durante a execução deste trabalho, que impuseram um consumo excessivo de tempo durante a implementação do modelo computacional, propõe-se para futuros avanços neste trabalho a avaliação das previsões em comparação com as observações com testes quantitativos, além da inspeção visual. Entre as diferentes hipóteses que poderiam ser testadas, propõe-se que testes de valores mínimos, máximos e médios fossem realizados para cada nível hierárquico do modelo Bayesiano, esclarecendo ainda mais a qualidade das previsões.

7 Conclusão

Com as devidas implementações deste projeto, foi possível demonstrar como as variáveis preditoras e/ou de controle e o modelo de inferência Bayesiano interagem em diferentes níveis de IDH para prever o número de mortos por COVID-19. De acordo com os resultados apresentados, a variável número total de casos parece ser fundamental para predizer o número de mortes, especialmente em localizações com níveis intermediários de IDH.

Pode-se constatar a paridade entre ambos os modelos analisados comprovando que independentemente da abordagem, os resultados tendem a ser próximos o suficiente para concluirmos que a predição se manteve coerente, e com base nos valores de predição, prior predictive check e posterior predictive check, entendemos que o modelo satisfaz o que era esperado para a abordagem, inferindo o numero de mortes por COVID-19 esperado dado os resultados coletados de diversos países.

Caso o leitor se interesse em reproduzir o modelo para outros dados ou abordando outros problemas característicos como o problema de morte por COVID-19, sugerimos a utilização do STAN sendo aplicado no R (linguagem de programação), evitando problemas de compatibilidade com o python dos quais encontramos nesta implementação.

Referências

[1] C. C. for Disease Control and Prevention, "The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (covid-19): China," *China CDC Weekly*, vol. b, 2020.

[2] S. Baloch, M. A. Baloch, T. Zheng, and X. Pei, "The coronavirus disease 2019 (COVID-19) pandemic," *The Tohoku Journal of Experimental Medicine*, vol. 250, no. 4, pp. 271–278, 2020.

[3] E. J. Williamson, A. J. Walker, K. Bhaskaran, S. Bacon, C. Bates, C. E. Morton, H. J. Curtis, A. Mehrkar, D. Evans, P. Inglesby, J. Cockburn, H. I. McDonald, B. MacKenna, L. Tomlinson, I. J. Douglas, C. T. Rentsch, R. Mathur, A. Y. S. Wong, R. Grieve, D. Harrison, H. Forbes, A. Schultze, R. Croker, J. Parry, F. Hester, S. Harper, R. Perera, S. J. W. Evans, L. Smeeth, and B. Goldacre, "Factors associated with COVID-19-related death using OpenSAFELY," *Nature*, vol. 584, pp. 430–436, July 2020.

[4] K. Tekalign, "Probable factors contributing to the fast spread of the novel coronavirus (COVID-19) in ethiopia," *Journal of Infectious Diseases and Epidemiology*, vol. 6, Oct. 2020.

[5] S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC Press, 2011.

[6] Stan Development Team, "RStan: the R interface to Stan," 2018. R package version 2.17.3.